

RESEARCH ARTICLE

A Novel Lung Nodule Accurate Segmentation of PET-CT Images Based on Convolutional Neural Network and Graph Model

XUNPENG XIA, AND RONGFU ZHANG, (Member, IEEE)

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

Corresponding author: Xunpeng Xia (191550053@st.usst.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 61971275 and Grant 81830052.


ABSTRACT Positron Emission Tomography and Computed Tomography (PET/CT) imaging could obtain functional metabolic feature information and anatomical localization information of the patient body. However, tumor segmentation in PET/CT images is significantly challenging for fusing of dual-modality characteristic information. In this work, we have proposed a novel deep learning-based graph model network which can automatically fuse dual-modality information for tumor area segmentation. Our method rationally utilizes the advantage of each imaging modality (PET: the superior contrast, CT: the superior spatial resolution). We formulate this task as a Conditional Random Field (CRF) based on multi-scale fusion and dual-modality co-segmentation of object image with a normalization term which balances the segmentation divergence between PET and CT. This mechanism considers that the spatial varying characteristics acquire different scales, which encode various feature information over different modalities. The ability of our method was evaluated to detect and segment tumor regions with different fusion approaches using a dataset of PET/CT clinical tumor images. The results illustrated that our method effectively integrates both PET and CT modalities information, deriving segmentation accuracy result of 0.86 in DSC and the sensitivity of 0.83, which is 3.61% improvement compared to the W-Net.

INDEX TERMS PET/CT images, graph model, deep learning, fusion learning.

I. INTRODUCTION

Positron emission tomography and Computed Tomography (PET/CT) imaging with 18F-fluorodeoxyglucose (FDG) have been widely used in cancer diagnosis. PET imaging extracts metabolic and functional information about organs from the human body. CT imaging extracts detailed anatomical high-resolution information of the human body. Compared with the CT modality, PET imaging can be implemented for an earlier diagnosis of the disease. According to the characteristics of inconsistent metabolic absorption rate of pathological and normal tissue, the pathological tissues appear as “high contrast” in PET images [1]. In the PET modality, the typically high contrast is presented between the malignant tumor and normal tissue [2]. However, PET

imaging is limited by imaging principles. The image which obtained has the disadvantage of low spatial resolution, which results in blurred lesion area boundary. In addition, a tumor area usually presents intensity distribution inhomogeneity in the PET modality. For these reasons, accurately delineating a tumor edge from a single PET modality is arduous. Compared with the PET modality, the CT modality provides high-resolution details of anatomical information from the human body. Under the same conditions, CT images have a higher spatial resolution detail than PET images. There are clear edges between the malignant tumors and peripheral normal tissue in the CT modality. Unfortunately, the intensity distribution of the normal soft tissues is usually similar to that of the tumor areas in the CT. Since it is formidable to segment the tumor regions from the surrounding tissues in CT images when a malignant tumor invades into the adjacent normal soft tissues. This situation often occurs in the detection of

The associate editor coordinating the review of this manuscript and approving it for publication was Shadi Alawneh .

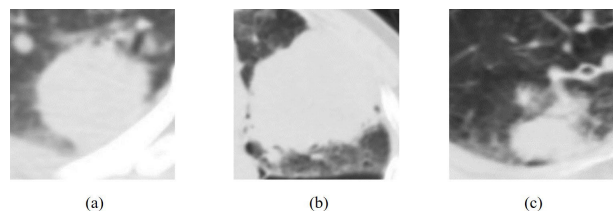


FIGURE 1. Tumor invade the chest wall.

lung cancer. A non-small cell lung cancer tumor can invade the chest wall or the thoracic vertebrae. Besides, CT images have complex interference information from the background. Complex background information can lead to arduous tumor segmentation (as Figure 1). Based on these reasons above, it is extremely difficult to obtain a clear boundary segmentation of tumors using a single modality.

Researches in recent years illustrate that the fusion strategy of multiple modalities can effectively improve the accuracy of diagnosis [3]. Multiple modalities imaging techniques, such as MRI/PET, MRI/SPECT, MRI/CT and PET/CT, have been applied to clinical trial [3], [4], [5]. Multiple modalities fusion techniques provide a method to obtain the information of focal area from biological and physical aspects. Fusion strategies are proposed to combine the complementary information from multi-modality images [6]. In the decade, an army of multiple-modality tumor segmentation approaches based on machine learning and level sets have been proposed. In the level set aspect, [7] used Jensen-Renyi divergence to ergodic update the level set contour and proposed a multimodality segmentation method based on geometric level contour. Reference [8] proposed a level set model based on multi-valued integration. This method can effectively aggregate multiple-modality and realize cross-modality data iteration of the level set. However, the level set algorithm has a high requirement on the area ratio between foreground object and background, minuscule object will result in segmentation failure. Although the level set has a strong anti-noise ability, the segmentation target would be lost in the image with a complex structure [9]. To address these issues, relevant kinds of literature [10] obtain images suitable for level set segmentation in advance through a multitude of pre-processing of images to be segmented. In the machine learning domain, the main research direction is the application of traditional feature engineering and probability graph models. For example, [11] comprehensively used watershed segmentation, the dynamic threshold to perform a multi-modality fusion of texture features of CT and metabolic features of PET and classifies the global model with support vector machine (SVM) as the classifier. References [12], [13], [14], and [15] similarly used metabolic and texture features engineering to merge multi-modality image and SVM classifier for the staging of lymphoma patients. Reference [16] formulates the problem of segmenting tumors from CT and PET modalities as Markov Random Field (MRF) with specific-modality energy terms for the characteristic of PET and CT. Reference [15] proposed a

random walk approach as an initial preprocessor to acquire an object original state. And the application of graph cutting method to segment pulmonary tumors on PET/CT modalities. However, conventional feature engineering methods have the disadvantages of difficult feature function setting and weak generalization ability.

With the development of deep learning networks, more and more researchers begin to apply deep learning frameworks so as to realize the fusion of multiple-modality [17]. A large number of experimental reaseaches have demonstrated the success of deep learning in the field of object segmentation [18], [19], image recognition [20], [21], [22], object detection [23], [24] and medical image processing [25]. Medical image processing mainly focuses on image processing of various modalities, such as CT image segmentation [26], [27], [28], standard-dose PET image estimation from low-dose PET/MRI [29] and multi-channel MRI image segmentation [30], [31], [32]. In the first MICCAI (Medical Image Computing and Computer Assisted Intervention) challenge on tumor segmentation with PET image, the application of convolutional neural network (CNN) won excellent results. The general deep learning network consists of multiple layers of processing units. Reasonable processing units can extract multi-scale representation data information from multi-modal medical images. High dimensional feature information can be extracted efficiently from complex structural data effectively by constructing a suitable multi-layers deep network model. [33] used a two-stage classification method to design a CNN network which can judge whether the candidate is a false positive sample. To obtain high and low-level features, many researchers have introduced V-Net. Reference [34] introduced first V-Net [35] is used to obtain CT image and the second V-Net is used to obtain pre-fused PET-CT image. Reference [36] also used V-Net for lung tumor segmentation. By enhancing the V-Net link mode, V-Net can be applied to the multi-branch paradigm. Similarly, some researches combine U-Net [37] with graph model to fuse PET modality and CT modality. Reference [14] trained U-Net for PET image and CT image portion respectively, fusing the multi-modality using a graph cut algorithm. Reference [9] proposed an improved model that integrated pixel intensity of PET with CT probability map from a CNN-derived to segment lung lesion. According to the researches on deep learning-based PET-CT multi-modality lung tumor segmentation focused on fusing image pixel intensity around the tumor. Yet, a single-pixel intensity integrating the PET-CT modalities ignores the relatives of PET and CT image features for tumors occurring in diverse anatomical locations. In addition, considering the relatives between multi-modality in the anatomical structure can significantly improve the segmentation accuracy.

In recent researches, graph-based dual-modality segmentation optimization obtains a army of attention in lesion segmentation/detection domain [38], [39], [40]. For enhancing integration of the complementary information from PET

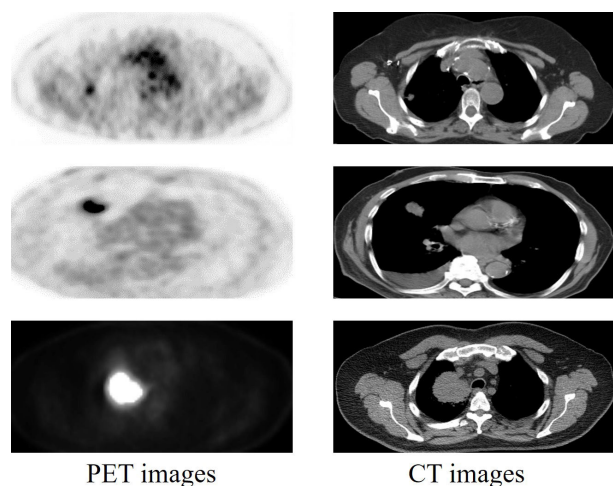


FIGURE 2. PET and CT images.

modality and CT modality images for pulmonary lesion segmentation, we propose an efficient graph-based network to strengthen each modality for target segmentation. The core idea of a graph-based algorithm is to transform the object task as the unary and pairwise energy minimization optimization problem, which can efficiently solve dual-modality features for target segmentation in the anatomical locations. First, in order to automatically obtain the feature information of each modality, we used CNN to learn the superior contrast metabolic information of PET and the superior spatial resolution information of CT images respectively and combine the features of the dual-modality for joint segmentation (as the Fig 2).

Then, the modality features of PET and CT images obtained by the superior are fused, which is formulated as a Conditional Random Field (CRF) probability estimation problem. We introduce a global co-fusion energy term into the target loss function for balancing the segmentation diverse between PET and CT images so that it can achieve lesion segmentation in dual-modality simultaneously. Through experiments, our novel network guarantees to achieve global optimization problems with the co-fusion energy term in a low-order polynomial formulation by computing each modality's maximum flow in the probability graph. After testing clinical data, the proposed method improves radiation therapy target definition and achieves 0.86 of the DSC on the co-segmentation.

II. METHOD

Figure 3 illustrates the framework architecture of our proposed approach. Our network comprises three main components: dual-modality encoder, multi-scale fusion component, and multi-modality reconstruction component. The main purpose of the dual-modality encoder (the green dotted box in Figure 3) is to derive each modality image features that are maximum relevant to the corresponding specific image modality. We use the Convolution Neural Network (CNN)

to obtain two-dimension slice image data from each modality. The multi-scale fusion component (the blue dotted box in Figure 3) utilizes the modality-specific features yielded by the encoders of the previous level to derive a series of different scales spatially diverse feature maps. In order to make use of multi-scale spatially feature maps reasonably, we construct state transition probability models between different scales across diverse multiple scales. Ultimately, the multi-modality reconstruction component (the red dotted box in Figure 3) integrates dual-modality-specific fused features by graph-based structure optimization to produce the final segmentation map. The crucial components are further described in detail in the next subsections.

A. DUAL-MODALITY ENCODER

The dual-modality encoder consists of two branches: one encoder for PET images and one separate encoder for CT images. The purpose of each modality encoder is to acquire the modality-specific features that are corresponding to the input modality data. According to research in recent years, CNN achieves high-precision object detection and segmentation in medical images [41]. Therefore, we use the cascade convolutional layers to build each modality encoder. As shown in Figure 3 each modality encoder comprises five blocks, each of which contains two convolution kernels and a pooling layer. Each level of block is connected to a side-output to provide feature maps for multi-scale fusion components.

Although the CNN has achieved a satisfactory result in the pulmonary probability map, a single CNN still has several limitations. First, a single CNN has a convolutional kernel with fixed receptive fields. Therefore, it yields coarse pixel-level resulting maps [42]. Second, a CNN absents fine-grained constraints, which would lead to the loss of texture details of output result images [43]. In a complex cascade structure, these limitations above make subtle changes in the feature information of the underlying layer potentially affecting the deeper convolution layer. During the training stage, this potential impact means that even tiny errors are magnified in the course of multiple iterations of training and multi-level convolution calculations. In order to avoid the accumulation of errors in each convolution layer, we used the side-output to output the feature distribution maps of the convolution block at each scale. Then, the feature graph output from each side-output is modeled as the probability graph of chain state transition. As shown in Figure 4, a CNN absents fine-grained constraints, which would lead to loss of texture details of output result images. Since we utilize PGM to model the side-outputs of each stage. We design a linear PGM encoder for dual-modality independently. The graph network of each modality contains six nodes and five corresponding edges. In the graph network, each node is composed of CNN feature maps of multi-scales. Each edge is a connection of two adjacent scales.

Finally, we analyze the inter-scale relationship between each scale and continuously compensate for the low-level

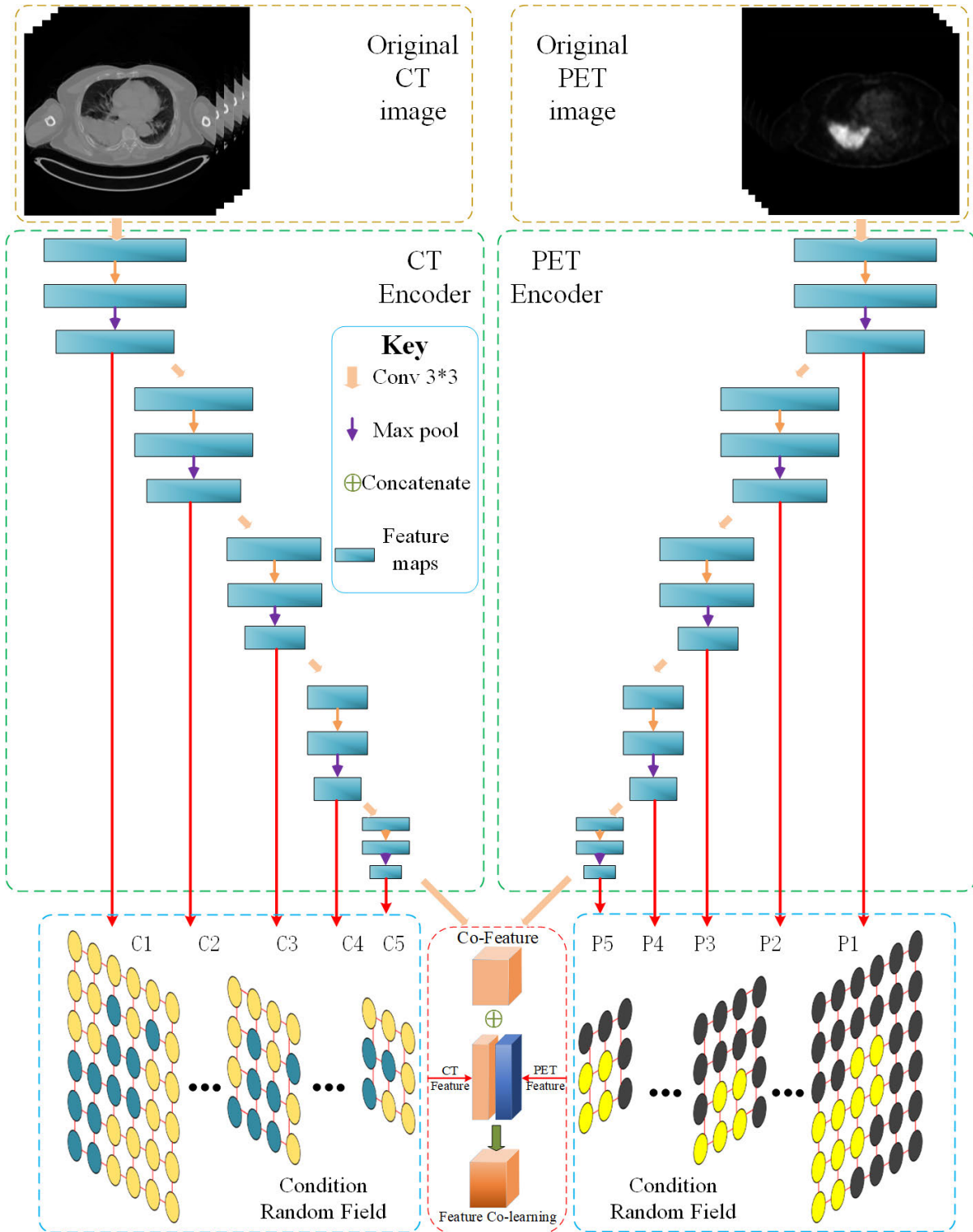


FIGURE 3. The architecture of our network.

details information lost due to multiple convolutional sampling. To effectively model the multi-scale output, we implement Conditional Random Field(CRF) to analyze the adjacent side-output state in the convolution component,

which is the major chain structure PGM. In the side-output, suppose W denotes the weight of each level of the convolutional layers, and N is defined as side-output layers in the feature sampling phase, where the corresponding weights are

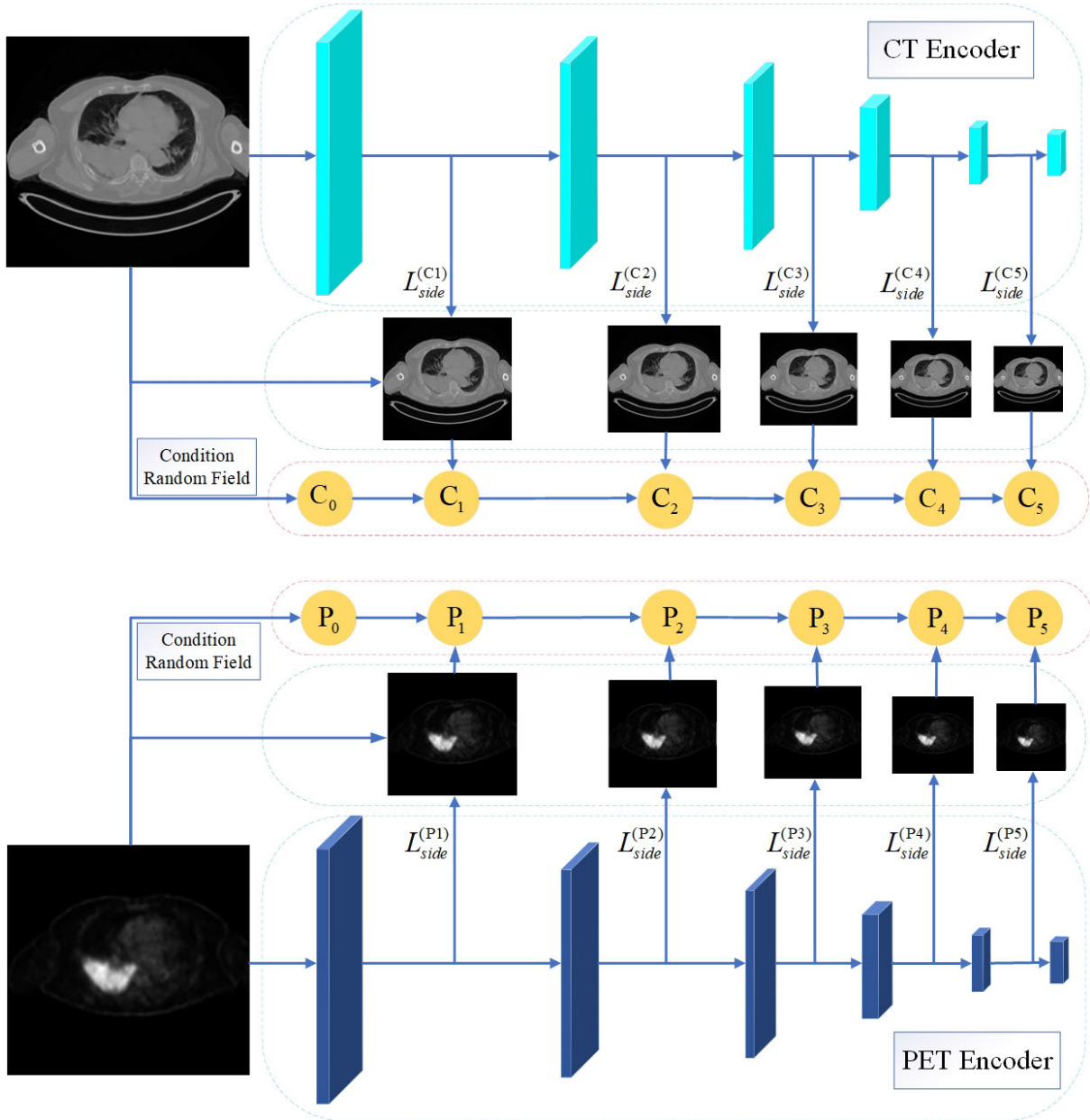


FIGURE 4. The linear chain CRF for dual-scale fusion.

defined as $w = (w^{(1)}, \dots, w^{(N)})$. The objective loss function of each side-output layer is as follows:

$$L_{Side}(W, w) = \sum_{n=1}^N \lambda_n L_s^{(n)}(W, w^{(n)}) \quad (1)$$

In which λ_n is the loss function linear combination weights of each side-output layer, and L_s defines the loss function of the prediction with ground-truth, which is quantified over all pixels of the training lung each modality image and corresponding lesion ground-truth.

The main purpose of using side-output is to obtain the characteristic map of each scale. The backbone of the whole

network is path-connected with each side-output. Therefore, the side-output layer parameters can continuously be updated according to back-propagation by the path of the weighted-fusion layer error propagation. After obtaining the feature maps of each scale, we use linear chain CRF for modeling to achieve deep supervision and multi-scale fusion. The following sections describe the CRF modeling process in detail.

B. MULTI-SCALE FUSION COMPONENT

Figure 4 shows the modeling and fusion process of multi-scale feature maps. We formulate the feature maps of each scale as the state transition graph of linear-chain CRF.

In the CRF stage, we mainly model the relationship of the unary and pairwise factors between adjacent scales. First, the conditional probability distribution $P(Y|x)$ is modeled as a CRF with the Gibbs distribution of:

$$P(Y = y|x) = \frac{1}{Z(x)} \exp(-E(V)) \quad (2)$$

where $E(V)$ is the energy function that measures the cost of unary potential and pairwise potential. We define $V = v_i$ as a labeling value over all pixels of the side-output image, with $v_i = 1$ for pulmonary nodules and $v_i = 0$ for normal tissue. The energy objective function of a label assignment V is given by:

$$E(V) = \sum_i \psi_u(v_i) + \sum_{i<j} \psi_p(v_i, v_j) \quad (3)$$

where $\psi_u(v_i)$ and $\psi_p(v_i, v_j)$ are the unary and pairwise terms respectively. The formula is as follows:

$$\psi_u = \frac{1}{N} \sum_{n=1}^N a_i^{(n)} \quad (4)$$

$$\psi_p(v_i, v_j) = w(v_i, v_j) \sum_{g=1}^G \lambda^g k^g(f_i, f_j) \quad (5)$$

$a_i^{(n)}$ is the value at pixel i in the edge prediction maps of side-output layer n . $w(v_i, v_j)$ is a trainable weight that coordinates the correlation intensity between two pairwise within the training stage. λ^g is the corresponding weight factor. k^g is the kernel of Gaussian applied on feature vectors, called transfer characteristics. The f_i and f_j are feature vectors of pixel i, j , respectively. They are derived from image features such as gray intensity values and spatial location coordinates. In this paper, we used the log-likelihood function form of conditional probability $P(Y = y|x)$.

$$\begin{aligned} L_s^{CRF}(\Lambda) &= \log P(Y = y|x) \\ &= \sum_{m=1}^M \sum_i \lambda^g k^g(a_i, f_i, f_j) - \log(Z_0) \end{aligned} \quad (6)$$

In our framework, we reformulate CRF as a Recurrent Neural Network(RNN) layer and can be implemented in an end-to-end framework.

To sum up, the objective function of the entire framework is:

$$L_{Total} = \operatorname{argmin}(L_{Side}(W, w) + L_s^{CRF}(\Lambda)) \quad (7)$$

where h is a layer parameter of CRF. L_{Side} and L_s^{CRF} are the loss functions of CNN layer and CRF stages respectively.

We implement the standard stochastic gradient descent to optimize the objective function. To comprehensively consider the actual situation of lung images, we selected five CNN blocks with side-output in the modeling stage. Pulmonary lesions in original medical images are different from the general object detection task in the visible light images. The general object detections contain rich semantic information,

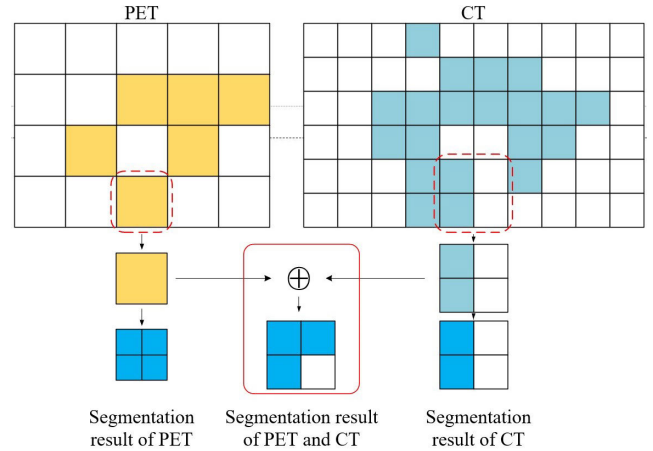


FIGURE 5. The PET/CT co-segmentation.

which allows the objective regions to be conserved in the high-level layers. In contrast, the medical image contains relatively single semantic information and a pulmonary lesion appears as a spot, which results in difficulty in obtaining responses in the high-level convolution layers. The low-level layers have a smaller size of the receptive field and reflect location information, while higher-level layers represent semantic information of a larger receptive scale.

The main function of the dual-modality encoder component is to extract the corresponding feature distribution map from PET and CT modalities. This process can not fuse the two modalities. The modality fusion part is mainly in the Multi-modality Reconstruction Component.

C. MULTI-MODALITY RECONSTRUCTION COMPONENT

The main function of the Multi-modality Reconstruction component is to learn the feature map information of each modality from the upper level and integrate the complementary information of the dual-modality. Figure 3 shows the fusion process of PET and CT images in the red dotted box. We formulate the task of co-fusion as the binary labeling of Conditional Random Field(CRF) on a probability map corresponding to the input CT and PET modality. Not only the framework attempts to simultaneously minimize the total CRF energy for both PET and CT modalities but also balances the segmentation result diverse between dual-modality. Finally, the output probability prediction map is reconstructed into a segmentation result image.

Figure 5 illustrates the co-segmentation process of PET and CT images. For two co-registered PET and CT images, the superior spatial resolution of CT and the superior intensity contrast of PET can achieve a more accurate segmentation result by simultaneously utilizing the information fusion method.

In our co-fusion problem, we set a discrete random variable (f_i, f_j) to represent each pixel value (i, j) of the input PET and CT images. Denote by V_{pet} and V_{ct} the set of the variable value corresponding to the pixels in the input PET and CT image respectively. In order to deal with

the resolution inconsistency between PET and CT images, we first implement the image registration algorithm to register the coordinate information of the CT and PET images. Then, the upsampling of PET images with low resolution guarantees that the resolution is consistent with that of CT images. Via image registration and upsampling, pixel coordinates between V_{pet} and V_{ct} of the dual-modality image can be kept in one-to-one correspondence with each other. Each label f in V_{pet} or V_{ct} gains the label value from the set of pixel labeling $N = 0, 1$ assigning that the pixel is in the object area ($f = 1$) or the normal tissue ($f = 0$). Ultimately, we compute an optimal pulmonary lesion segmentation in the PET or CT image to minimize the corresponding CRF energy by implementing Boykov and Kolmogorov's graph cuts approach. However, the single-use of graph cut method can not obtain information from other modalities. To realize the fusion of PET and CT modalities, we introduce a co-segmentation parameter to combine the two modalities.

We set up the third set of co-segmentation parameters V_{P-C} to correlate with a pair of the corresponding pixel (V_{pet}, V_{ct}) in the PET and CT modality. The co-segmentation parameter V_{P-C} is utilized for incorporating the energy balance of the segmentation diverseness between the dual-modality. Scilicet, when the pixel labeling pairs (f_i, f_j) are the same as the corresponding pixel coordinates (i, j), then no penalty is enforced; Otherwise, we balance the divergence from the dual-modality. Moreover, the degree of divergence between PET and CT modalities may be different in the results of final segmentation. Thus, we need to optimize for varying degrees of difference in the co-segmentation of pixel label pairs.

The task of PET-CT co-fusion is to minimum optimize the following energy loss function (Eq(8)).

$$L_{PET-CT} = E_P(V_{pet}) + E_C(V_{ct}) + E_{P-C}(F_{P-C}) \quad (8)$$

where $E_P(V_{pet})$ and $E_C(V_{ct})$ are the CRF energy functions for the PET and CT modality, respectively. The energy term of co-segmentation $E_{P-C}(F_{P-C})$ is utilized for balancing the segmentation divergence from the PET and CT. The co-segmentation energy term $E_{P-C}(F_{P-C})$ integrates the high contrast metabolism information of PET and the superior spatial resolution of CT to associate results of the PET segmentation and CT segmentation as a joint process. The various energy functions are described in detail below.

1) THE CRF SEGMENTATION ENERGY ON THE PET MODALITY

According to the state transition relationship of CRF, we model the neighborhood system on the input image data. Denote N_p is the neighborhood regions of the pixel I_p in the PET modality. Since, the CRF energy term in the PET modality consists of a current state term $d_i(f_i)$ and a smoothness term $w_{i-1,i}$, as follows:

$$E_P(V_{pet}) = \sum_{i \in I_p} d_i(f_i) + \sum_{i \in N_p} w_{i-1,i}(f_{i-1}, f_i) \quad (9)$$

The current state $d_i(f_i)$ is the likelihood that executes individual penalty for assigning a pixel label f_i (eg, the lesion or the normal) to the corresponding pixel i . The smoothness term $w_{i-1,i}(f_{i-1}, f_i)$, meaning the interaction potential between adjacent pixels (f_{i-1}, f_i) [44], estimates the loss of assigning diverse pixel labels to two adjacent pixels f_{i-1} and f_i in the set of N_p .

$$w_{i-1,i}(f_{i-1}, f_i) = \begin{cases} \alpha(i-1, i), & \text{if } f_{i-1} \neq f_i, \\ 0, & \text{if } f_{i-1} = f_i. \end{cases} \quad (10)$$

where $\alpha(i-1, i)$ is the smoothness value calculated from adjacent pixels when f_{i-1} and f_i are unequal. In the PET modality, we obtain an optimal segmentation concerning the energy function $E_P(V_{pet})$ by applying the graph cuts approach.

2) THE CRF SEGMENTATION ENERGY ON THE CT MODALITY
Denote N_c is the neighborhood regions of the pixel I_c in the CT modality. Since the CRF energy term in the CT modality consists of a current state term $d_j(f_j)$ and a smoothness term $w_{j-1,j}$, as follows:

$$E_C(V_{ct}) = \sum_{j \in I_c} d_j(f_j) + \sum_{j \in N_c} w_{j-1,j}(f_{j-1}, f_j) \quad (11)$$

$$w_{j-1,j}(f_{j-1}, f_j) = \begin{cases} \alpha(j-1, j), & \text{if } f_{j-1} \neq f_j, \\ 0, & \text{if } f_{j-1} = f_j. \end{cases} \quad (12)$$

In the CT modality, the current state term and smoothness term are the same as in the PET modality. We also use the graph cut algorithm to optimize the energy function.

3) THE CO-SEGMENTATION ENERGY TERM

To balance the divergence between the dual-modality, co-segmentation energy term $E_{p-c}(F_{P-C})$ is set to coordinate the segmentation diverse between the PET and the CT. Each variable value $f_{(i,j)} \in F_{P-C}$ relevant with a pair of corresponding pixels (i, j) in I_p and I_c assigns a pixel label from the labeling set $Y = \{0, 1\}$. When $f_{(i,j)} = 1$, the corresponding pixels pair of i and j are assigned as foreground (lesion region) labeling; Otherwise, the corresponding pixels pair of i and j are assigned as background (normal tissue). When f_i, f_j and $f_{(i,j)}$ are inconsistent, we use the $\eta_{i,j}(f_i, f_{(i,j)})$ to resolve the divergence of the f_i and f_j . Based on this condition $\eta_{i,j}$ can further resolve the difference based on the obvious feature from the PET and CT modalities. The function is described in detail in the following formula (13):

$$\eta_{i,j}(f_i, f_{(i,j)}, f_j) = \begin{cases} \delta_1(i, j), & \text{if } f_i \neq f_j, f_{(i,j)} = f_i, \\ \delta_2(i, j), & \text{if } f_i \neq f_j, f_{(i,j)} = f_j, \\ 0, & \text{if } f_i = f_{(i,j)} = f_j. \end{cases} \quad (13)$$

where $\delta_1(i, j)$ and $\delta_2(i, j)$ are the term to balance the segmentation divergence between corresponding pixel pair i and j . Summary the co-segmentation energy term between the PET I_p and the CT I_c is defined as follows:

$$E_{p-c}(F_{P-C}) = \sum_{i \in I_p, j \in I_c} \eta_{i,j}(f_i, f_{(i,j)}, f_j) \quad (14)$$

III. THE GRAPH MODEL OPTIMIZATION PROCESS

In this section, we present a graph-based method for solving the co-segmentation based on the PET-CT task. The resolution of co-segmentation achieves global network optimal with respect to the object energy function L_{PET-CT} defined in Equal(8). To solve the problem, we construct a graph model $G = (v, E)$ and solve it by calculating the minimum-cost cut in the low-order polynomial term. Figure 6 shows the general process of modality fusion between PET and CT by graph model.

Each pair of variables f_i and f_j are mapped to one pair node in v_{pet} and v_{ct} . In the graph G , a pair of i and j also denote the corresponding a pair node of pixel i and j in (I_p, I_c) . Moreover, a graph model node $\phi_{i,j} \in v_{p-c}$ is introduced for each variable value $f_{(i,j)} \in F_{p-c}$. Thus our purpose is to formulate the co-segmentation task as a minimum-cost cut problem. The minimum-cost cut problem consists of two terminal nodes, a source node s and a sink node t . All of the other vertices v_{p-c} have to be connected to these two vertices to form part of the edge set. The set of graph model $v = \{s, t\} \cup v_{ct} \cup v_{pet} \cup v_{p-c}$. The graph cuts method of Boykov and Kolmogorov contains two types of nodes and edges linking methods. One link method is between vertices corresponding to each pixel in the input image, defined as *N-links*. Another link is the connection between the vertex corresponding to each pixel in the input image and the two terminal vertices, defined as *T-links*. However, the standard graph linking method cannot model the input image of the two modalities. In order to model dual-modality, we introduce the three linking models: *T-links* and *N-links* model the sub-node of each modality from v_{pet} and v_{ct} respectively, and additional *D-links* connects the image data over the modality v_{p-c} . The following subsections introduce the three types of links in detail.

A. T-LINKS

The graph cut method is similar to the pixel labeling assignment problem. Our purpose is to minimize the energy function by cutting an optimal boundary between the object and the background. Therefore, to search for this optimal boundary, we need to consider the region of pixels on the current (i.e. Current state term) and the effect from adjacent regions of pixel (i.e. Smoothness term). We introduce *T-links* to integrate the current state term of the CRF segmentation energy. For each node i in v_{pet} from PET modality, we defined it as an edge start s to i with the edge cost function of $d_i(f_i = 1)$ and an edge start i to t with the edge cost function of $d_i(f_i = 0)$. Denote F and B are lesion region and normal tissue sets respectively. The problem of label assignment for each pixel is calculated by Gaussian Mixture Model (GMM) to reflect the probability intensity distribution of each pixel in I_p . The experiments prove that the fusion effect of GMM is the high point in this paper. Then the negative log-likelihoods of each pixel for assigning $F = 1$ or $B = 0$ obtained by GMM are used on $d_i(f_i)$. Defining the intensity value of each pixel i in I_p

as E_i . The pixel values of d_i are calculated for in the following formulation:

$$d_i(f_i) = \begin{cases} -\lambda_1 \ln P(E_i|F), & \text{if } f_i = F, \\ -\lambda_2 \ln P(E_i|B), & \text{if } f_i = B. \end{cases} \quad (15)$$

Similarly, we utilize *T-links* for each node j in v_{ct} from CT modality.

B. N-LINKS

We introduce *N-links* to measure the effect of each corresponding pixel intensity on its surrounding pixels (i.e., the smoothness terms $w_{i-1,i}$ in Equal(10)). For each pixel i in the PET modality I_p , we measure the elements in the neighborhood of the current pixel i by *N-links*. First, N_p is defined as the neighborhood set of pixel i . The setting of N_p range of the neighboring sets mainly considers experimental accuracy and computational cost. Then, we add two connection methods to link each adjacent pixel of pixel i . Note that the connection direction of each pair of adjacent pixels is bi-directional, i.e. one starts the node $i \in v_{pet}$ to the adjacent node in N_p and the other in the opposite direction start the adjacent node in N_p to node i . The cost value of each link path is $\alpha(i-1, i)$:

$$\alpha(i-1, i) = \begin{cases} \lambda_3 e^{(-\theta_1 \|E_{i-1} - E_i\|)}, & (i-1, i) \in N_p \\ 0, & \text{Otherwise.} \end{cases} \quad (16)$$

In the same way, the *N-links* are introduced for the sub-node set v_{ct} with the neighboring set N_c on the CT modality.

C. D-LINKS

D-links are utilized for measuring the segmentation divergence between the PET and CT modality. Similarly $i \in v_{pet}$ and $j \in v_{ct}$, we introduce a novel node $\Phi_{i,j} \in v_{p-c}$. Each node $\Phi_{i,j}$ from the v_{p-c} is connected to the corresponding node of PET and CT modality. For example in the PET modality, we set two links between i and $\Phi_{i,j}$. One link start i to $\Phi_{i,j}$ and the other one start $\Phi_{i,j}$ to i , each link with a cost function of $\phi_1(i, j)$. Those types of *D-links* are used to penalize the divergence case where $f_i \neq f_j$, but $f_{i,j} = f_j$. In the same way, two links between j and $\Phi_{i,j}$ are set with each of cost $\phi_2(i, j)$ on the CT modality. Those types of *D-links* are used to penalize the divergence case that $f_i \neq f_j$, but $f_{i,j} = f_i$. Both cost function $\phi_1(i, j)$ and $\phi_2(i, j)$ are calculated by below formulation:

$$\Phi(i, j) = \begin{cases} \phi_1(i, j) = \lambda_4 e^{(-\theta_2 \|E_i - E_j\|)} \\ \phi_2(i, j) = \lambda_5 e^{(-\theta_3 \|E_i - E_j\|)}. \end{cases} \quad (17)$$

Thus we complete the construction of the graph model $G = (v, E)$ from the PET and CT modality. Figure 7 illustrates a pixel sample construction of the graph model. In this work, we set the 8-neighboring pixels system for dual-modality N_p and N_c . In addition that, higher-order clique potential learning can acquire more complex interactions information of conditional random variables. However, the computation cost for learning the clique potential increases exponentially

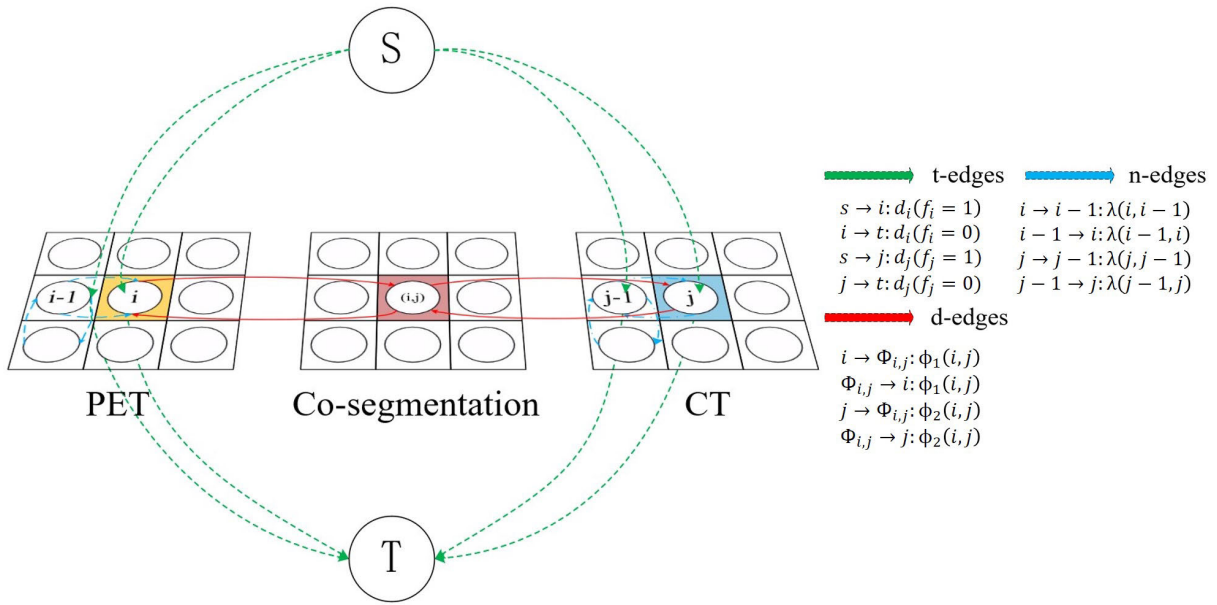


FIGURE 6. The construction of co-segmentation graph.

with the range of the clique and result in a difficult problem of energy minimization.

Based on the structure of graph model $G = (v, E)$ above, we further demonstrate the minimum-cost cut. The core idea of the minimum graph cut problem is to binary divide graph by $s - t$ cut C and obtain the minimum weight. The graph cut C divides all nodes in the graph into two disjoint subsets S and T by cutting the edges of the graph model, where the source s is in subset S and the terminal point t is in subset T . Each edge $e \in E$ in the graph model is defined a non-negative weight w_e . A path cut is a subset of edges $C \subseteq E$ such that the terminals points separated on the induced graph $G(C) = \langle v, E|C \rangle$. Via normal combinatorial optimization, we combined statistics for each cutting path to calculate the sum value of the overall path cost, the following formula:

$$|C| = \sum_{e \in C} w_e. \quad (18)$$

Graph cut formalism is suitable for pixel-level segmentation tasks. The total nodes of the graph $G = (v, E)$ represent pixels from the dual-modality and the edges represent adjacent relevance between the pixels.

Figure 7 illustrates partitioning corresponds to the segmentation of underlying image pixels. Through built into the edge weights, a minimum cost cut path yields a segmentation result that is an optimal solution in terms of properties. Therefore, the optimal co-segmentation solution of PET and CT modality can be acquired by computing a minimum $s - t$ cut in G .

IV. EXPERIMENTAL SETTINGS

In this section, we show the parameters setting of our fusion network on dual-modality in detail. Via several experiments

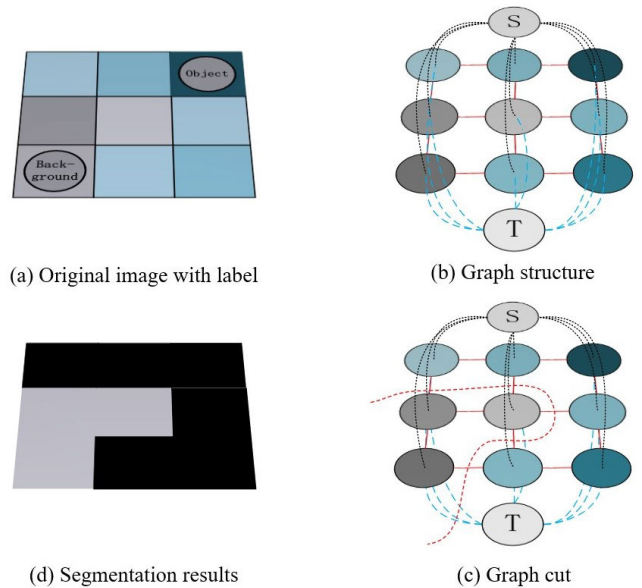


FIGURE 7. Diagram of graph cut algorithm.

verification, we demonstrate the advantages of our proposed network in PET-CT co-segmentation.

A. PREPARATION DATASETS

The proposed network was validated on the clinical medical dataset with 135 non-small cell lung carcinoma(NSCLC) patients(Collected from The Cancer Imaging Archive(TCIA) [45]). We randomly selected 88 PET/CT images to train our network, and the remaining 47 PET/CT images were used as the test dataset. To further verify the generalization ability of the proposed method, a partition of clinical data from [9] were used for validation.

Each pair of PET and CT images has been registered by the special hardware on the PET/CT scanner. The PET scanner's transaxial resolution increased from 4.6mm to 10mm from the center at a radius of 200mm. The axial resolution of PET increased from 3.5mm at the center to 7.8mm at a radius of 200mm [46], [47]. For the PET modality images acquisition, the PET images were reconstructed by the Ordered Subset Expectation Maximization(OSEM) method with 6 subsets and 2 iterations. The dimensionality of each PET slice reconstructed image is 128*128. For the CT modality images, the resolution of each CT slice reconstructed is 512*512.

The ground truth of each pair of the PET/CT images in the datasets was manual annotation by an experienced radiologist. With the existing lung lesion segmentation approaches, we defined a rectangular region of interest(ROI) for each PET/CT slice, which enclosed the whole lesion region. Our method and all comparison baseline methods were implemented in the ground truth.

B. COMPARISON METHODS

To systematically measure the performance of multi-modality information fusion to lung lesion segmentation, we compared our proposed method with the conventional methods(without deep learning), segmentation method based on deep learning respectively. The conventional methods include thresholding and level-set segmentation [48]. The deep learning methods include single modality such as FCN [49] and dual-modality such as V-Net [35], W-Net [50] and 3D-UNet+GC [51], [52]. Besides, we compared our method with some classical segmentation approaches on the PET modality. For example, thresholding algorithms-Otsu automatic thresholding and a graph theory method-Graph Cuts(GC) [52].

C. PARAMETER SETTING AND IMPLEMENTATION

Our proposed method was implemented with Python3.6 on a standard Linux server with a tesla P100 arithmetic processor. The graph model optimization tool selects the maximum-flow library [53]. For the PET/CT image registration, we utilized Elastix [54] tools to register it to the corresponding CT image. After dual-modality registering, we obtained PET/CT data with the same pixel resolution and one-to-one corresponding pixel coordinates of two modality images. In our experiments, we set the parameter as follows: The current state term coefficients: $\lambda_1 = \lambda_2 = 0.9$. The smoothness term coefficients, we set $\lambda_3 = 25$ and $\theta_1 = \theta_2 = \theta_3 = 1.2$. The co-segmentation term coefficients: $\lambda_4 = 25$ and $\lambda_5 = 1.1$. The learning rate of our network was 0.003. In addition that, the selection of fixed parameters is set via several rounds of experimental tuning.

D. EVALUATION METRIC

The segmentation methods performance was evaluated by calculating the Dice Similarity Coefficient(DSC), positive predictive value(PPV), classification error(CE), sensitivity(SE), and volume error(VE). SE, PPV, and DSC

calculate the similarity(value of spatial overlap) between the segmented lung lesion volume S_A and the ground-truth volume S_G :

$$SE = \frac{|S_A \cap S_G|}{|S_G|}. \quad (19)$$

$$PPV = \frac{|S_A \cap S_G|}{|S_A|}. \quad (20)$$

$$DSC(S_A, S_G) = \frac{2|S_A \cap S_G|}{|S_A + S_G|}. \quad (21)$$

These three evaluation metrics above are ranged from [0, 1](0:without spatial overlap, 1:perfect spatial overlap). According to recent research on the PET/CT tumor segmentation [55], we can calculate the accuracy *Score* by unifying PPV and SE, as the following formulation:

$$Score = 0.5(PPV + SE). \quad (22)$$

Compared with DSC, SE and PPV, CE, and VE measure the area difference and spatial location bias between the segmentation tumor pixel and the ground-truth volume [56].

$$CE(S_A, S_G) = \frac{abs(|S_A| - |S_G|)}{|S_G|}. \quad (23)$$

$$VE(S_{FP}, S_{FN}, S_G) = \frac{(|S_{FP}| - |S_{FN}|)}{|S_G|}. \quad (24)$$

where S_{FP} defines the number of false-positive samples, S_{FN} denotes the number of false-negative samples. The smaller value of CE and VE means a more accurate segmentation result.

V. EXPERIMENTAL RESULTS

In the training stage of the network, We show the training loss value of the entirety model and the weight convergence of each side-output. Figure 8 illustrates our proposed method training result and the baseline methods training results. The baseline methods include 50% Threshold, W-Net and the 3D-UNet+GC. From this figure, we observe that the loss value of the 50% Threshold network in the training initial phase, is up to 0.512. However, deep learning-based methods have lower loss value in the initial stage. W-Net, 3D-UNet+GC and our proposed method reach 0.188, 0.106, and 0.066 respectively. On the hand, this improvement benefits from the pre-training of the convolution network, which substantially shortens the training period. On the other hand, the introduction of PGM filters irrelevant feature subsets to a certain extent and improves training efficiency. In the final phase, the proposed method also has a low loss value of 0.044, which is better than other methods.

We also extracted all weight parameters of each side-output layer in the initial phase, 20 epoch, 40 epoch, and final phase of the training process for visualization, which is shown in Figure 9. In the initial phase, the distribution map of each *weight* is scattered, especially outside of the image. The central region of the image produces multiple secondary peaks with the extension of sides. In the 20-40 epoch stage, the scattered peaks on the sides of the image gradually weaken

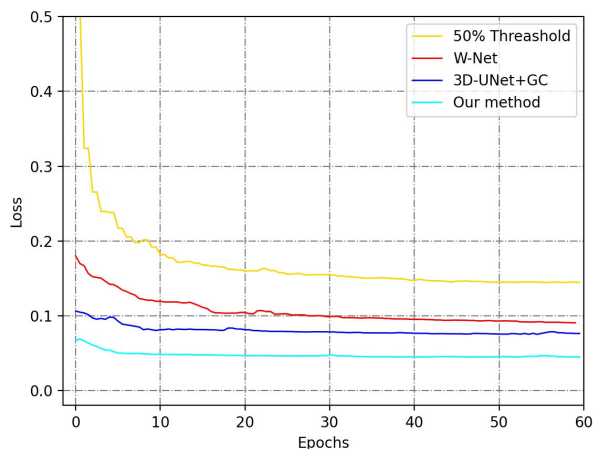


FIGURE 8. Comparison of our propose method with baseline on the model training stage performance.

or disappear. At the same time, the secondary peak in the central region tends to be flat and increasingly converges to the center. In the final phase, the scattered peaks on the sides have completely disappeared, while the secondary peak in the central region absolutely converges to the center. On the whole, the weight values in the side-output trend gradually converge after the four stages, and the whole detection network also completes the training.

VI. PERFORMANCE VALIDATION ON TUOMR SEGMENTATION

In this section, we mainly illustrate the performance of our network and baseline methods (FCN-CT, FVM-PET, Threshold, Level set, V-Net, 3D-UNet+GC, W-Net) on PET/CT tumor segmentation. First, we compare the probability map performance between our method and baseline on PET/CT segmentation.

Figure 10 displays three slices images of the isolated lesion from three different patient images data in the first col. The first image shows areas of non-small cell lung cancer, while others show regions of microscopic nodules. We selected three representative probability map feature extraction methods, which named 50% Threshold, W-NET and our method. In general, all three methods have achieved excellent results for feature extraction of large tumor areas in the first image. However, the 50% Threshold method is least effective in the other two microscopic nodules with serious identification errors. Although the features extracted by W-Net accurately delineated the lesion areas, it lost an army of texture information and the probability of image contrast is lower. In contrast, the multi-scale method added in this paper can accurately extract the contour of the lesion core region while retaining the texture information in the region completely.

Figure 11 shows three slices of lesion images that overlap with the surrounding tissue. Both the 50% Threshold and FCN-CT methods confused the lesion area with the pleura. In contrast, our method and W-Net can effectively distinguish

the lesion and pleural areas. Compared with the W-Net model, our method can retain texture information of the lesion area, providing a basis for further segmentation tasks.

Although the probability maps of tumor regions could not accurately delineate the tumor outline, probability maps could roughly reflect the relationship between the tumor and its surrounding normal tissues. The probability maps are also the basis for accurate delineating, for the intensity of the response has a high similarity to the tumor. This indicated that the probability maps obtained by our proposed network could effectively depict and distinguish the tumor and background region. This is paramount for our method to further fuse the PET and CT information for accurate segmentation of the tumor area.

Figure 12 illustrates the segmentation results of FVM-PET and FCN-CT in a single modality. In figure 12, the black curve is the ground-truth by manual delineating, and the blue curve is the segmentation region outlined by the two methods respectively. We can find that the segmentation recognition of a single modality has confusion between the lesion region and normal tissue. In the PET modality, the trachea (Figure 12(c)) is divided into lesion areas and the necrotic areas in the tumor center (Figure 12(g)) is divided into normal tissue by FVM-PET. In CT modality, although the FCN-CT method can effectively segment normal tracheal tissue, it is prone to interference from surrounding groups with similar contrast. The main reason for this defect is the lack of multimodal complementary information.

Figure 13 displays the segmentation results of W-Net, 3D-UNet+GC and our method (the pink curve is ground truth; the red curve is segmentation results of our method). In the images with complex background (in the Figure 13(a)(b)), 3D-UNet+GC presents a relatively serious segmentation failure (as the blue curve). In the central necrotic areas, the dual-modality contradictory information balancing failed and the edge smoothness is insufficient. In the Figure 13(c)(d), the segmentation results obtained by 3D-UNet+GC are more accurate and correctly segment the trachea region. However, the fineness of the delineation is still insufficient and the visual effect is rough. The segmentation results of W-Net has higher fineness and smoother curve (as the green curve). Though the segmentation of necrotic areas has also been improved, there are deviations in tracheal region division in Figure 13(d).

Figure 14 is a series of visual comparison results of the segmentation for a PET-CT image slice with nodules within the lung tissue. This figure depicts that our proposed network could effectively segment the pulmonary nodules within the mediastinum, although the resulting image has slightly deviated from the ground-truth. In contrast, none of the other methods can effectively segment lung nodules in the mediastinum.

Compared with the above two methods, our method is the closest to the curve drawn by ground-truth (pink curve) and has the highest overlap area. In the visual, our method is significantly superior to the W-Net and 3D-UNet+GC

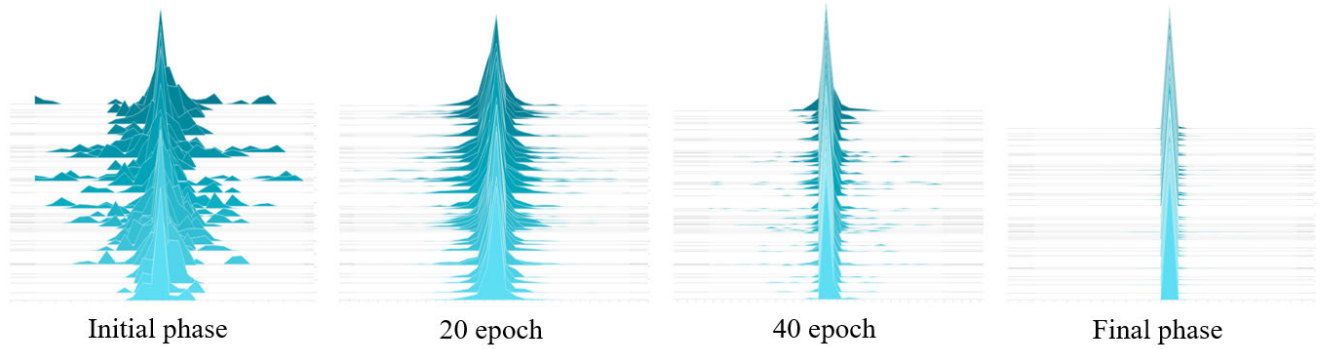


FIGURE 9. Convergence of side-output in different training phases.

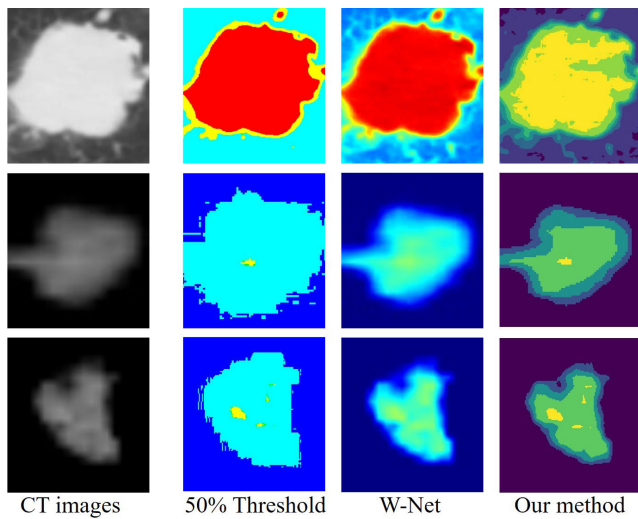


FIGURE 10. Visualization probability map of our proposed method with 50% Threshold and W-Net method on the isolated lesion.

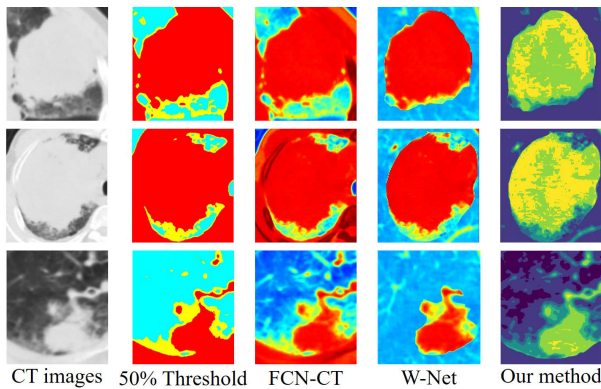


FIGURE 11. Visualization probability map of our proposed method with 50% Threshold, FCN-CT and W-Net method in the case of adhesion to surrounding normal tissue.

approaches on segmentation accuracy and test stability. A probable factor is that we used the probability map fusion by graph cut algorithm for the multi-modality segmentation while the other deep learning-based methods used complex

dual-modality images register. In another word, this demonstrates the superior capacity of our designed method to finely delineate the tumor region and the normal tissue using a graph model.

In table 1, we list the detailed numerical evaluation value(DSC, SE, PPV, CE, and VE). Obviously, our method is superior to the other methods which used the single-modality(PET or CT) image information. The major factor for the inferior single-modality segmentation performance of the PCN-CT method is that the intensity inhomogeneity from the PET modality is left out of consideration. FVM-PET, which only uses PET modality, generally has higher indexes than FCN-CT. The main reason is that the strong intensity contrast of PET images provides a better segmentation basis. Similarly, we also compare the proposed method and several multi-modality tumor segmentation algorithms with other deep learning-based fusion strategies. The proposed method has a higher DSC and Score value of segmentation results and is more stable than the W-Net and 3D-UNet+GC.

In the table 2, we also counted the detailed numerical evaluation value(Train, Test, IoU, and Inter). Similarly to other baseline approaches, the performance of our proposed network is close to the ground-truth, even though a significantly larger number of dataset has been learned. Advantageously, our method does not rely on complicatedly heavy pre-processing stages and allows to segment nodules of all textures and sizes without the need to define specific parameters.

In conclusion, our method combining the advantages of deep learning multi-scale features and specially designed graph models to construct segmentation networks is helpful for more accurate segmentation of tumor regions from PET/CT dual-mode.

VII. DISCUSSION

PET/CT modality images have been widely implemented in clinical medical practice. Generally, the imaging of the PET modality has a high-intensity contrast with a low spatial resolution. The tumor boundary is blurred on the morphological and tumor regions may have intensity inhomogeneity in the PET modality. Relatively, the characteristic of CT imaging

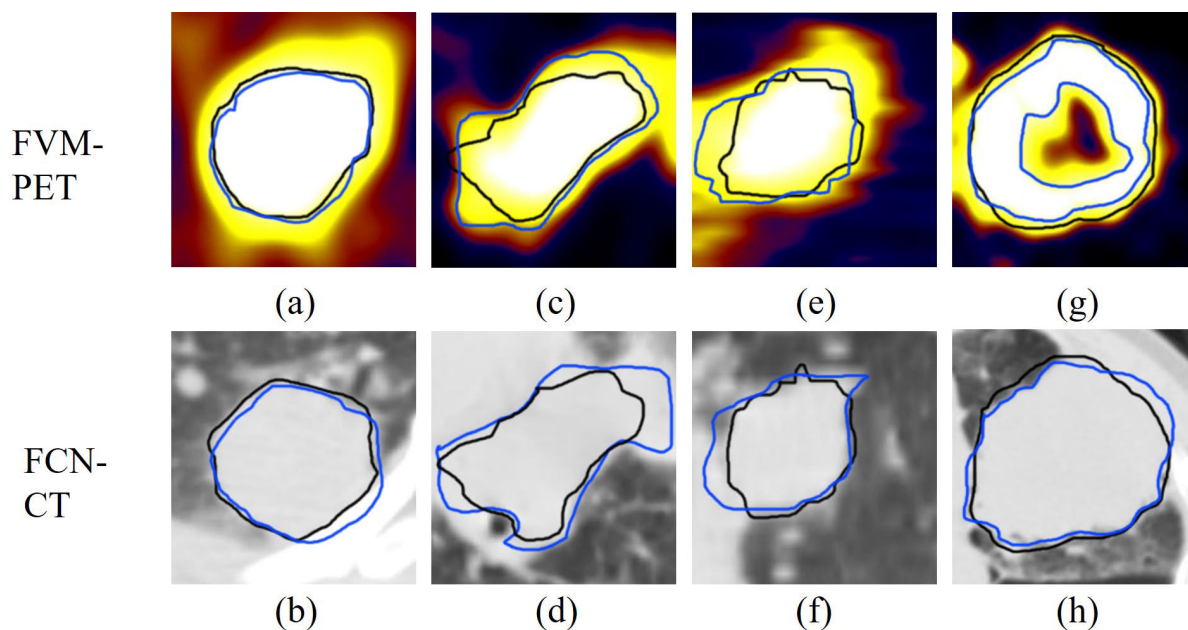


FIGURE 12. Segmentation result of FVM-PET and FCN-CT in the single modality.

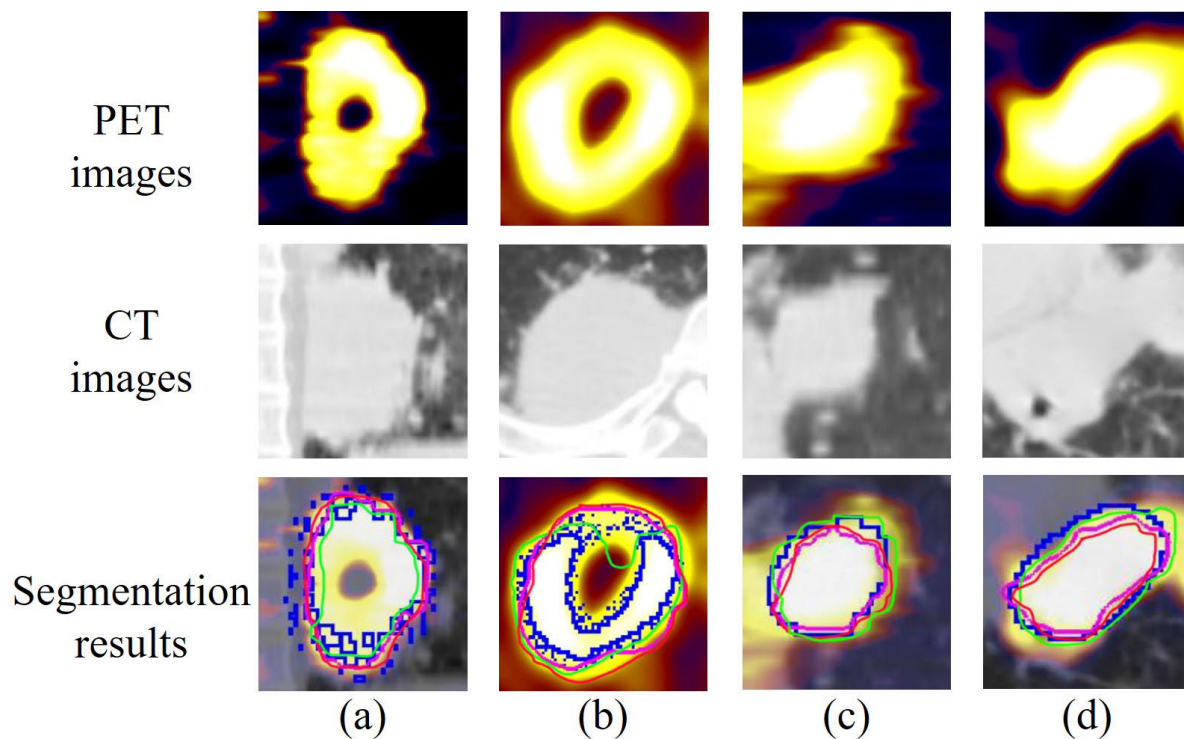


FIGURE 13. Segmentation result of W-Net, 3D-UNet+GC and our method in the dual-modality.

has superior spatial resolution with a low-intensity contrast between the lesion region and neighboring normal tissues. Integrating the advantages of the dual-modality can enhance the tumor segmentation relative accuracy. Nowadays,

multi-modality tumor co-segmentation by PET/CT still has some shortcomings. Since the complementary information from the PET and CT images could be contradictory. In this paper, we proposed a novel network to integrate

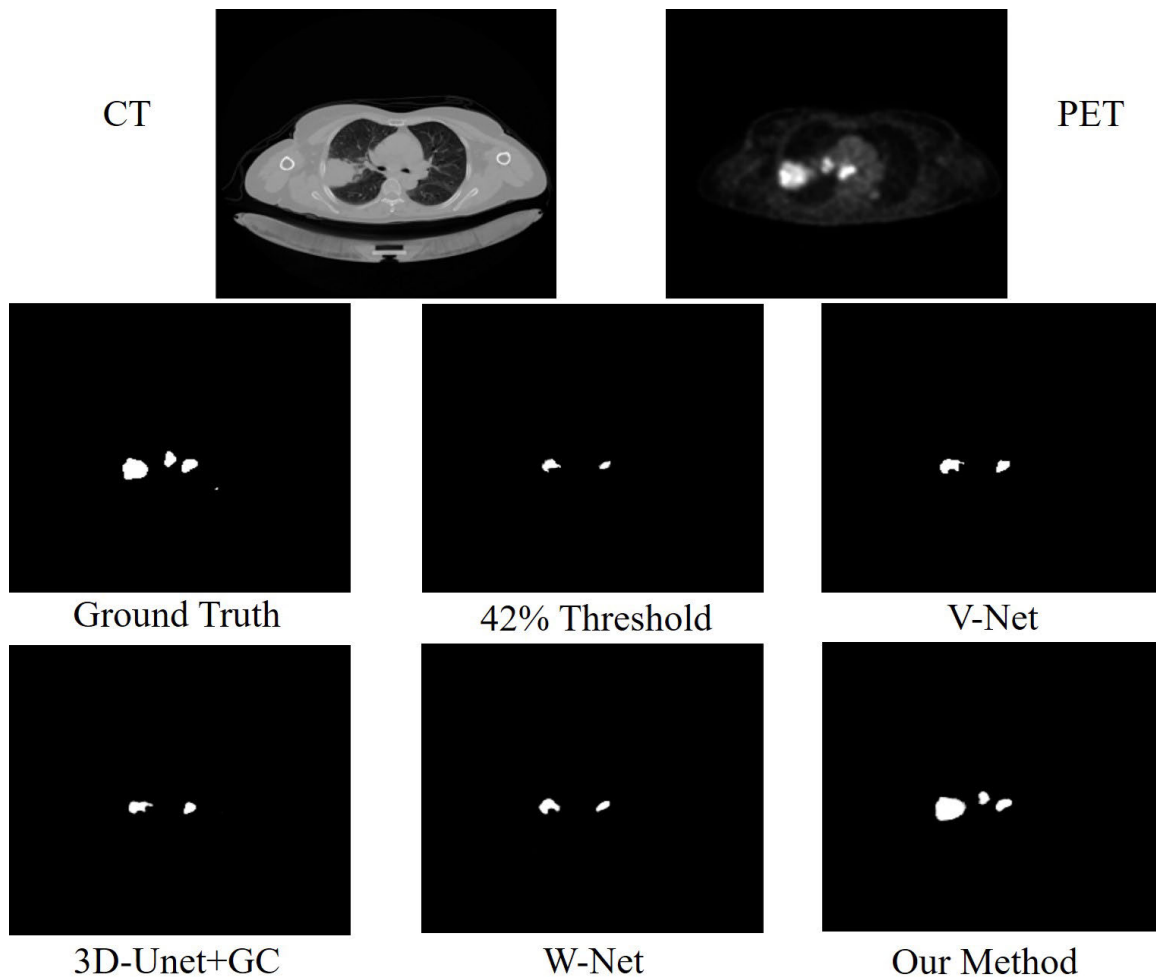


FIGURE 14. Visual results comparison of the segmentation obtained by our proposed method compared to the four baselines and the ground-truth(GT).

TABLE 1. The mean value of DSC, SE, PPV, Score, VE and CE of the segmentation results of different segmentation algorithms.

Methods	DSC	SE	PPV	Score	VE	CE
FCN-CT	0.76	0.82	0.74	0.78	0.28	0.53
FVM-PET	0.82	0.82	0.85	0.84	0.25	0.38
42% Threshold	0.70	0.61	0.90	0.76	0.38	0.47
50% Threshold	0.58	0.45	0.94	0.69	0.54	0.51
Level Set(JR)	0.78	0.75	0.86	0.80	0.26	0.42
V-Net	0.73	0.87	0.67	0.77	0.48	0.46
3D-UNet+GC	0.82	0.86	0.81	0.84	0.22	0.37
W-Net	0.83	0.87	0.82	0.84	0.17	0.36
Our	0.86	0.83	0.91	0.87	0.25	0.28

the PET/CT complementary information for tumor area segmentation.

On the PET modality, several normal tissue areas neighboring the tumor could have a highly similar intensity response. Since these normal tissue areas could lead to incorrectness segmentation results as tumors by conventional segmentation methods. To improve the segmentation accuracy of PET, the proposed method correctly segments areas as normal tissue

with the constraint of a prior tumor probability graph model from CT.

On the CT modality, the tumor areas express various sizes and shapes in complex forms. The tumor also has a similar intensity response with its surrounding normal tissue region. It is arduous for conventional approaches to describe the tumor on a single CT image. The state-of-the-art deep learning-based network is used to handle the complex forms

TABLE 2. The mean value of Train, Test, IoU and Inter of the segmentation results of different segmentation algorithms.

Methods	Train	Test	IoU	Inter
FCN-CT	508	43	0.76±0.13	0.75±0.11
FVM-PET	653	63	0.73±0.07	0.76±0.05
42% Threshold	NA	34	0.62±0.15	0.65±0.06
50% Threshold	NA	32	0.53±0.05	0.54±0.07
Level Set(JR)	NA	47	0.82±0.12	0.78±0.09
V-Net	63	517	0.69±0.16	0.77±0.12
3D-UNet+GC	300	64	0.79±0.04	0.81±0.06
W-Net	1404	1404	0.82±0.02	0.84±0.01
Our	1593	1593	0.87±0.03	0.85±0.06

of CT images. Generally, the excellent capability of deep learning is based on the learning of a large multitude of datasets with ground-truth labels. However, in the medical image processing domain, especially on PET/CT modalities, there are problems with obtaining ground-truth labels data are arduous and acquiring amount of learning data. In our work, only a few batch patients were applied to our deep learning network for learning. First, the tumor region is effectively separated by a probability map from the complex background. Then graph model further fuses the PET and CT modality information for more accurate segmentation of tumor regions. In addition, we use the probability graph model and side-output mechanism to construct a whole set of multi-scale feature learning networks. The complementary information of high-level and low-level is used to compensate for the information loss of multiple convolution sampling.

In the fusion dual-modality information, we built a novel graph model construction to enhance the strength of each of the PET and CT modalities for tumor regions segmentation. Inspired by information provided by other modalities in which tumors are simultaneously segmented from PET and CT images. We design the weight coefficients to balance the divergence and contradictory information from each modality. For measuring the inter-relationship between the pixels at each coordinate position and surrounding areas, the graph cut algorithm is used to refine co-segmentation results. Through the improvements above, our co-segmentation method can successfully integrate both information from PET and CT images. Normally, we assume that dual-modality images could be perfectly registered, i.e. each pixel one-to-one correspondence between PET and CT. However, image registration bias is hard to void and would seriously affect the fusion results. In this paper, the algorithm of dual-modality running on a low-order polynomial term by optimizing graph G has weak robustness against registration bias [16]. Consequently, the high-order terms are introduced to experiment with the robustness of the fusion segmentation model in our future work.

VIII. CONCLUSION

To further improve the accuracy of the co-segmentation model, we proposed a novel supervised deep learning

network for fusing complementary feature information from dual-modality image data. Our method leverages CNN to derive a series of spatial diverse fusion probability maps from the dual-modality specific features. Then, quantifying the relevance of each modality pixel across varying spatial locations by the graph model. Our achievement from lesion region detection and segmentation experiments on PET/CT non-small lung cancer images demonstrated that our proposed method significantly enhanced (improved 3.61% than W-Net on DSC) than several baseline deep learning-based approaches for dual-modality image tasks. Experiments illustrate that our conceptual method which has a specific graph model mechanism component to derive fusion probability maps, which could be a serviceable technique for medical image analysis applications especially requires integrating complementary feature information from diverse image modalities.

REFERENCES

- [1] C. A. Mathis, "A lipophilic thioflavin-T derivative for positron emission tomography (PET) imaging of amyloid in brain," *ChemInform*, vol. 12, no. 3, pp. 295–298, Feb. 2002.
- [2] C. S. Voskuilen, E. J. van Gennep, S. M. H. Einerhand, E. Vegt, M. L. Donswijk, A. Bruining, H. G. van der Poel, S. Horenblas, K. Hendricksen, B. W. G. van Rhijn, and L. S. Mertens, "Staging ^{18}F -fluorodeoxyglucose positron emission tomography/computed tomography changes treatment recommendation in invasive bladder cancer," *Eur. Urol. Oncol.*, vol. 5, no. 3, pp. 366–369, Jun. 2022.
- [3] L. Beyer, A. Gosewisch, S. Lindner, F. Vltter, and H. Ilhan, "Dosimetry and optimal scan time of [^{18}F]SiTATE-PET/CT in patients with neuroendocrine tumours," *Eur. J. Nucl. Med. Mol. Imag.*, vol. 48, no. 11, pp. 3571–3581, Oct. 2021.
- [4] A. Shoeibi, M. Khodatars, M. Jafari, P. Moridian, M. Rezaei, R. Alizadehsani, F. Khozeimeh, J. M. Gorriz, J. Heras, M. Panahiazar, S. Nahavandi, and U. R. Acharya, "Applications of deep learning techniques for automated multiple sclerosis detection using magnetic resonance imaging: A review," *Comput. Biol. Med.*, vol. 136, Sep. 2021, Art. no. 104697.
- [5] C. J. Laing, T. Tobias, D. I. Rosenblum, W. L. Banker, L. Tseng, and S. W. Tamarkin, "Acute gastrointestinal bleeding: Emerging role of multidetector CT angiography and review of current imaging techniques," *RadioGraphics*, vol. 27, no. 4, pp. 1055–1070, Jul. 2007.
- [6] R. Liu, J. Liu, Z. Jiang, X. Fan, and Z. Luo, "A bilevel integrated model with data-driven layer ensemble for multi-modality image fusion," *IEEE Trans. Image Process.*, vol. 30, pp. 1261–1274, 2021.
- [7] D. Markel, I. E. Naqa, C. Freeman, and M. Vallières, "SU-E-J-110: A novel level set active contour algorithm for multimodality joint segmentation/registration using the Jensen-Rényi divergence," *Med. Phys.*, vol. 39, no. 6Part7, p. 3678, Jun. 2012.

- [8] D. Yang, A. Apte, D. Khullar, S. Mutic, J. Zheng, J. D. Bradley, P. Grigsby, and J. O. Deasy, "Concurrent multimodality image segmentation by active contours for radiotherapy treatment planning," *Med. Phys.*, vol. 34, no. 12, pp. 4738–4749, 2007.
- [9] L. Li, X. Zhao, W. Lu, and S. Tan, "Deep learning for variational multimodality tumor segmentation in PET/CT," *Neurocomputing*, vol. 392, pp. 277–295, Jun. 2020.
- [10] D. Markel, C. Caldwell, H. Alasti, H. Soliman, Y. Ung, J. Lee, and A. Sun, "Automatic segmentation of lung carcinoma using 3D texture features in 18-FDG PET/CT," *Int. J. Mol. Imag.*, vol. 2013, pp. 1–13, Feb. 2013.
- [11] J. Zhao, G. Ji, Y. Qiang, X. Han, B. Pei, and Z. Shi, "A new method of detecting pulmonary nodules with PET/CT based on an improved watershed algorithm," *PLoS ONE*, vol. 10, no. 4, Apr. 2015, Art. no. e0123694.
- [12] C. Lartizien, M. Rogez, E. Niaf, and F. Ricard, "Computer-aided staging of lymphoma patients with FDG PET/CT imaging based on textural information," *IEEE J. Biomed. Health Informat.*, vol. 18, no. 3, pp. 946–955, May 2014.
- [13] Y. Song, W. Cai, H. Huang, X. Wang, Y. Zhou, M. J. Fulham, and D. D. Feng, "Lesion detection and characterization with context driven approximation in thoracic FDG PET-CT images of NSCLC studies," *IEEE Trans. Med. Imag.*, vol. 33, no. 2, pp. 408–421, Feb. 2014.
- [14] Z. Zhong, Y. Kim, K. Plichta, B. G. Allen, L. Zhou, J. Buatti, and X. Wu, "Simultaneous cosegmentation of tumors in PET-CT images using deep fully convolutional networks," *Med. Phys.*, vol. 46, no. 2, pp. 619–633, 2018.
- [15] W. Ju, D. Xiang, B. Zhang, L. Wang, I. Kopriva, and X. Chen, "Random walk and graph cut for co-segmentation of lung tumor on PET-CT images," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5854–5867, Dec. 2015.
- [16] D. Han, J. Bayouth, S. Qi, A. Taurani, and X. Wu, "Globally optimal tumor segmentation in PET-CT images: A graph-based co-segmentation method," in *Proc. 22nd Int. Conf. Inf. Process. Med. Imag.*, 2011, pp. 245–256.
- [17] Y. H. Bhosale and K. S. Patnaik, "Application of deep learning techniques in diagnosis of COVID-19 (coronavirus): A systematic review," *Neural Process. Lett.*, vol. 16, pp. 1–53, Sep. 2022, doi: [10.1007/s11063-022-11023-0](https://doi.org/10.1007/s11063-022-11023-0).
- [18] S. Wang, G. Sun, B. Zheng, and Y. Du, "A crop image segmentation and extraction algorithm based on mask RCNN," *Entropy*, vol. 23, no. 9, p. 1160, Sep. 2021.
- [19] Y. H. Bhosale and K. S. Patnaik, "IoT deployable lightweight deep learning application for COVID-19 detection with lung diseases using RaspberryPi," in *Proc. Int. Conf. IoT Blockchain Technol. (ICIBT)*, May 2022, pp. 1–6, doi: [10.1109/ICIBT52874.2022.9807725](https://doi.org/10.1109/ICIBT52874.2022.9807725).
- [20] A. Mahendran and A. Vedaldi, "Visualizing deep convolutional neural networks using natural pre-images," *Int. J. Comput. Vis.*, vol. 120, no. 3, pp. 233–255, Dec. 2016.
- [21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
- [22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
- [23] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
- [24] Y. H. Bhosale and K. S. Patnaik, "PulDi-COVID: Chronic obstructive pulmonary (lung) diseases with COVID-19 classification using ensemble deep convolutional neural network from chest X-ray images to minimize severity and mortality rates," *Biomed. Signal Process. Control*, vol. 81, Mar. 2023, Art. no. 104445.
- [25] S. Suganyadevi, V. Seethalakshmi, and K. Balasamy, "A review on deep learning in medical image analysis," *Int. J. Multimedia Inf. Retr.*, vol. 11, no. 1, pp. 19–38, 2022.
- [26] P. F. Christ, M. Elshaer, F. Ettliger, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, M. Armbruster, F. Hofmann, and M. D'Anastasi, "Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2016, pp. 415–423.
- [27] H. R. Roth, A. Farag, L. Le, and E. B. Turkbey, "Deep convolutional networks for pancreas segmentation in CT imaging," in *Proc. SPIE*, vol. 9413, pp. 378–385, Mar. 2015.
- [28] Q. Dou, H. Chen, Y. Jin, L. Yu, J. Qin, and P. A. Heng, "3D deeply supervised network for automatic liver segmentation from CT volumes," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2016, pp. 149–157.
- [29] L. Xiang, Y. Qiao, D. Nie, L. An, W. Lin, Q. Wang, and D. Shen, "Deep auto-context convolutional neural networks for standard-dose PET image estimation from low-dose PET/MRI," *Neurocomputing*, vol. 267, pp. 406–416, Dec. 2017.
- [30] K. Kamnitsas, C. Ledig, V. F. J. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Med. Image Anal.*, vol. 36, pp. 61–78, Feb. 2017.
- [31] X. Zhao, Y. Wu, G. Song, Z. Li, Y. Zhang, and Y. Fan, "A deep learning model integrating FCNNs and CRFs for brain tumor segmentation," *Med. Image Anal.*, vol. 43, pp. 98–111, Jan. 2018.
- [32] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P. M. Jodoin, and H. Larochelle, "Brain tumor segmentation with deep neural networks," *Med. Image Anal.*, vol. 35, pp. 18–31, Jan. 2017.
- [33] A. Teramoto, H. Fujita, O. Yamamuro, and T. Tamaki, "Automated detection of pulmonary nodules in PET/CT images: Ensemble false-positive reduction using a convolutional neural network technique," *Med. Phys.*, vol. 43, pp. 2821–2827, May 2016.
- [34] L. Xu, G. Tetteh, J. Lipkova, Y. Zhao, H. Li, P. Christ, M. Piraud, A. Buck, K. Shi, and B. H. Menze, "Automated whole-body bone lesion detection for multiple myeloma on 68Ga-pentixafor PET/CT imaging using deep learning methods," *Contrast Media Mol. Imag.*, vol. 2018, pp. 1–11, Jan. 2018.
- [35] X. Luo, W. Zeng, W. Fan, S. Zheng, and Y. Chen, "Towards cascaded V-Net for automatic accurate kidney segmentation from abdominal CT images," in *Proc. SPIE*, vol. 11596, pp. 345–351, Feb. 2021.
- [36] X. Zhao, L. Li, W. Lu, and S. Tan, "Tumor co-segmentation in PET/CT using multi-modality fully convolutional neural network," *Phys. Med. Biol.*, vol. 64, no. 1, Dec. 2018, Art. no. 015011.
- [37] J. Wang, X. Zhang, P. Lv, L. Zhou, and H. Wang, "Ear-U-Net: EfficientNet and attention-based residual U-Net for automatic liver segmentation in CT," in *Proc. Comput. Vis. Pattern Recognit.*, 2021, pp. 10–14.
- [38] D. Han, J. E. Bayouth, S. Bhatia, M. Sonka, and X. Wu, "Motion artifact reduction in 4D helical CT: Graph-based structure alignment," in *Proc. Int. MICCAI Workshop Med. Comput. Vis.*, 2011, pp. 63–73.
- [39] Q. Song, M. Chen, J. Bai, M. Sonka, and X. Wu, "Surface-region context in optimal multi-object graph-based segmentation: Robust delineation of pulmonary tumors," in *Proc. Biennial Int. Conf. Inf. Process. Med. Imag.*, 2011, pp. 61–72.
- [40] Q. Song, X. Wu, Y. Liu, M. Sonka, and M. Garvin, "Simultaneous searching of globally optimal interacting surfaces with shape priors," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2879–2886.
- [41] M. K. Hasan, L. Calvet, N. Rabbani, and A. Bartoli, "Detection, segmentation, and 3D pose estimation of surgical tools using convolutional neural networks and algebraic geometry," *Med. Image Anal.*, vol. 70, May 2021, Art. no. 101994.
- [42] H. Fu, Y. Xu, S. Lin, D. W. K. Wong, and J. Liu, "DeepVessel: Retinal vessel segmentation via deep learning and conditional random field," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2016, pp. 132–139.
- [43] Z. Xiao, N. Du, L. Geng, F. Zhang, J. Wu, and Y. Liu, "Multi-scale heterogeneous 3D CNN for false-positive reduction in pulmonary nodule detection, based on chest CT images," *Appl. Sci.*, vol. 9, no. 16, p. 3261, Aug. 2019.
- [44] M. Beheshti and W. C. Liew, "Image segmentation based on graph-cut models and probabilistic graphical models: A comparative study," in *Machine Learning and Cybernetics (Communications in Computer and Information Science)*. Berlin, Germany: Springer, 2015.
- [45] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, and F. Prior, "The cancer imaging archive (TCIA): Maintaining and operating a public information repository," *J. Digit. Imag.*, vol. 26, no. 6, pp. 1045–1057, Dec. 2013.
- [46] W. Wadsak and M. Mitterhauser, "Basics and principles of radiopharmaceuticals for PET/CT," *Eur. J. Radiol.*, vol. 73, no. 3, pp. 461–469, Mar. 2010.
- [47] J. L. Humm, A. Rosenfeld, and A. Del Guerra, "From PET detectors to PET scanners," *Eur. J. Nucl. Med. Mol. Imag.*, vol. 30, no. 11, pp. 1574–1597, Nov. 2003.
- [48] V. Rajinikanth, "Appraisal of breast ultrasound image using Shannon's thresholding and level-set segmentation," in *Proc. Prog. Comput., Anal. Netw.*, 2019, pp. 621–630.

- [49] J. Cheng, Z. Ren, Q. Zhang, X. Gao, and F. Hao, "Cross-modality compensation convolutional neural networks for RGB-D action recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 3, pp. 1498–1509, Mar. 2022.
- [50] H. Zhang, M. Wang, F. Wang, G. Yang, Y. Zhang, J. Jia, and S. Wang, "A novel squeeze-and-excitation W-Net for 2D and 3D building change detection with multi-source and multi-feature remote sensing data," *Remote Sens.*, vol. 13, no. 3, p. 440, Jan. 2021.
- [51] Z. Zhong, Y. Kim, L. Zhou, K. Plichta, B. Allen, J. Buatti, and X. Wu, "3D fully convolutional networks for co-segmentation of tumors on PET-CT images," in *Proc. IEEE 15th Int. Symp. Biomed. Imag.*, Apr. 2018, pp. 228–231.
- [52] Z. Zhong, Y. Kim, K. Plichta, B. G. Allen, L. Zhou, J. Buatti, and X. Wu, "Simultaneous co-segmentation of tumors in PET-CT images using deep fully convolutional networks," *Med. Phys.*, vol. 46, no. 2, pp. 619–633, 2019.
- [53] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1124–1137, Sep. 2004.
- [54] H. J. Kuijff, C. Tax, L. K. Zaanen, W. H. Bouvy, J. D. Bresser, A. Leemans, M. A. Viergever, G. J. Biessels, and K. L. Vincken, "The added value of diffusion tensor imaging for automated white matter hyperintensity segmentation," in *Computational Diffusion MRI (Mathematics and Visualization)*. Cham, Switzerland: Springer, 2014, pp. 45–53.
- [55] M. Hatt, B. Laurent, A. Ouahabi, H. Fayad, S. Tan, L. Li, W. Lu, V. Jaouen, C. Tauber, and J. Czakov, "The first MICCAI challenge on PET tumor segmentation," *Med. Image Anal.*, vol. 44, pp. 177–195, Feb. 2018.
- [56] L. Li, J. Wang, W. Lu, and S. Tan, "Simultaneous tumor segmentation, image restoration, and blur kernel estimation in PET using multiple regularizations," *Comput. Vis. Image Understand.*, vol. 155, pp. 173–194, Feb. 2017.



XUNPENG XIA received the B.S. degree in electrical engineering and the M.S. degree in mechanical engineering from Anhui Jianzhu University, Anhui, China, in 2015 and 2018, respectively. He is currently pursuing the Ph.D. degree with the School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, China. His research interests include image processing, medical image, and multi-modality fusion.



RONGFU ZHANG (Member, IEEE) received the B.S. degree in physics from Fuyang Normal University, in 1995, the master's degree in optical instruments from the University of Shanghai for Science and Technology, in 1998, and the Ph.D. degree in communication and information system from Shanghai Jiao Tong University, in 2004.

He was a Lecturer, in 2001, and an Associate Professor, in 2004. In 2005, he visited The University of Arizona. He was the Deputy Director and the Director of the Teaching and Research Section. Since 1998, he has been a Teacher with the University of Shanghai for Science and Technology. He has presided over four projects of national major instruments special project of the Ministry of Science and Technology and the Shanghai Automobile Industry Fund and Education Commission. He has published more than 20 academic articles (five in SCI and ten in EI). He has co-edited three textbooks and one translated book.

...